DOCUMENT RESUME

ED 080 588 TM 003 105

AUTHOR	Lindley, Dennis V.
TITLE	Multiple Regression in a Two-Way Layout.
INSTITUTION	American Coll. Testing Program, Iowa City, Iowa. Research and Development Div.
SPONS AGENCY	Office of Education (DHEW), Washington, D.C.
REPORT NO	ACT-TB-9
PUB DATE	Sep 72
GRANT	OEG-0-72-0711
NOTE	15p.
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	*Bayesian Statistics; *Mathematical Applications; *Mathematical Models; *Multiple Regression Analysis; *Predictor Variables; Statistical Studies; Technical Reports

ABSTRACT

This paper discusses Bayesian m-group regression where the groups are arranged in a two-way layout into m rows and n columns, there still being a regression of y on the x's within each group. The mathematical model is then provided as applied to the case where the rows correspond to high schools and the columns to colleges: the predictor variables might be the performances in various course area taken while at high school and the random variable the first-year performance at college. Then "cell" (i,j) of the two-way table will contain data on the performances of those subjects who passed from high school i to college j. Eighteen equations are given. (DB)

ACT TECHNICAL BULLETIN NO. 9

MULTIPLE REGRESSION IN A TWO-WAY LAYOUT

by

Dennis V. Lindley

The Research and Development Division

The American College Testing Program

P. O. Box 168, Iowa City, Iowa 52240

September, 1972

MULTIPLE REGRESSION IN A TWO-WAY LAYOUT[1]

by

Dennis V. Lindley

University College London

Suppose that in each of a number, m, of groups a random variable y has a linear regression on a set of predictor variables $x_1, x_2, \ldots, x_p$. A Bayesian approach to the estimation of the m regression lines, valid under certain assumptions of exchangeability, has been given by Lindley (1969, 1970). The theory has been described as part of the Bayesian analysis of the general linear model by Lindley and Smith (1972)--see, in particular, Section 3.2. The basic ideas have been extensively developed and put into an operational form by Jackson, Novick, and Thayer (1971) under the name of m-group regression and implemented in a major application by Novick, Jones, and Cole (1972). Technical details concerning the computer program are described by Jones and Novick (1972).

The present paper extends these ideas to the case where the groups, now m times n in all, are arranged in a two-way layout into m rows and n columns, these still being a regression of y on the x's within each group. The sort of application we have in mind is where the rows correspond to high schools and the columns to colleges: the predictor variables might be the performances in various course areas taken while at high school and the random variable the first-year performance at college. Then "cell" (i, j) of the two-way table will contain data on the performances of those subjects who passed from high school i to college j . In the case of m-group regression, the estimation

[1]The research reported herein was performed pursuant to Grant No. OEG-0-72-0711 with the Office of Education, U.S. Department of Health, Education, and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

of any single regression can be improved by using all the data and not just that from the group under consideration. Our aim is to effect similar improvements in the case of the two-way layout.

The situation where there are no predictor variables is a non-trivial special case. There, in our example, we would merely have the college score, and the familiar analysis-of-variance type of situation arises in which we try to separate out the variation into components due to schools (rows) and colleges (columns). This has been discussed by Lindley (1972). It turns out that the regression case is a straighforward multivariate generalization of the special situation. Essentially, all that happens is that the scalar equations for the special case become vector and matrix results in the general problem. The reader is, therefore, advised to read the simpler, earlier paper first. Having understood the basic ideas in the scalar context, the difficulties in the present, multivariate case should be substantially reduced.

The data are supposed to be generated by a model in which

$$E(y_{ijk}) = \sum_{s=1}^{p} \theta_{ijs} x_{ijks} \tag{1}$$

for $i = 1, 2, \ldots, m$; $j = 1, 2, \ldots, n$; $k = 1, 2, \ldots, r_{ij}$. There, y is the random variable having linear regression on p variables $x_1, x_2, \ldots, x_p$. The suffix i refers to the row, j to the column, s to the predictor variable, and k to the replicate number within a cell. Hence, θ_{ijs} is the regression coefficient of y on x_s in the group that is in the i^{th} row and j^{th} column.

It will further be supposed that the distribution of the y-values, conditional on the x-values and the regression coefficients, is normal; that they are all independent and have constant variance σ^2 . It is possible to relax this last condition and have the variances possibly different from cell to cell. This was done in the earlier papers referred to, but the work of Jackson, et al. (1971) suggests that the practical effect of such a generalization is minimal. Since it complicates the algebra, which is already formidable enough, we have not dealt with the heteroscedastic case here. The reader who wishes to study it will find the ideas needed discussed in Appendix 4 of Lindley (1972). This paper will subsequently be referred to as L .

In most applications, one of the predictor variables, say x_1, will be a constant, say one, to allow for a constant term in (1). If so, that equation may be written

$$E(y_{ijk}) = \theta_{ij1} + \sum_{s=2}^{p} \theta_{ijs} x_{ijks} \quad .$$

If $p = 1$, we have simply $E(y_{ijk}) = \theta_{ij}$, the third suffix for θ being redundant, and the situation is exactly that discussed in L . The prior structure assigned to the mn values θ_{ij} was to suppose

$$\theta_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij}$$

where $\alpha_i \sim N(0, \sigma_a^2)$, $\beta_j \sim N(0, \sigma_b^2)$, $\gamma_{ij} \sim N(0, \sigma_c^2)$, all these distributions being independent, and the prior knowledge of μ to be vague. The reasons given for this assumption were stated at length in L, but essentially, it follows from supposed exchangeability between rows and between columns plus assumptions of additivity and normality.

Now, we can equally write (1) as

$$E(y_{ijk}) = \underset{\sim}{\theta}_{ij}^{T} \underset{\sim}{x}_{ijk} \qquad (2)$$

where $\underset{\sim}{\theta}_{ij}$ is a vector given by

$$\underset{\sim}{\theta}_{ij}^{T} = (\theta_{ij1}, \theta_{ij2}, \ldots, \theta_{ijp}) \; ,$$

the vector of regression coefficients from cell (i, j), and $\underset{\sim}{x}_{ijk}$ a similar vector with typical element x_{ijks} . We make the same assumption of prior structure about the <u>vector</u> $\underset{\sim}{\theta}_{ij}$ that we made about the <u>scalar</u> θ_{ij} . Specifically, we assume

$$\underset{\sim}{\theta}_{ij} = \underset{\sim}{\mu} + \underset{\sim}{\alpha}_i + \underset{\sim}{\beta}_j + \underset{\sim}{\gamma}_{ij} \tag{3}$$

where $\underset{\sim}{\alpha}_i \sim N_p(\underset{\sim}{0}, \underset{\sim}{\Sigma}_a)$, $\underset{\sim}{\beta}_j \sim N_p(\underset{\sim}{0}, \underset{\sim}{\Sigma}_b)$, $\underset{\sim}{\gamma}_{ij} \sim N_p(\underset{\sim}{0}, \underset{\sim}{\Sigma}_c)$, all these distributions being independent and the prior knowledge of $\underset{\sim}{\mu}$ being vague. (There, N_p means multivariate, p-dimensional normal, and the $\underset{\sim}{\Sigma}$'s are the dispersion matrices.) Essentially, this means that the set of regression coefficients in cell (i, j) has possible effects ($\underset{\sim}{\alpha}_i$) due to the row, to the column ($\underset{\sim}{\beta}_j$), and to an interaction between row and column ($\underset{\sim}{\gamma}_{ij}$). The remarks in L about the restrictive nature of this assumption and the necessity of checking its reasonable validity before using the methods developed with it as a basis, apply with even more force in this multivariate case.

The form (3) implies the following covariance structure for the vectors $\underset{\sim}{\theta}_{ij}$:

$$\text{cov}(\underset{\sim}{\theta}_{ij}, \underset{\sim}{\theta}_{rs}^T) = 0 , \qquad i \neq r, j \neq s , \tag{4a}$$

$$\text{cov}(\underset{\sim}{\theta}_{ij}, \underset{\sim}{\theta}_{is}^T) = \underset{\sim}{\Sigma}_a, \qquad j \neq s , \tag{4b}$$

$$\text{cov}(\underset{\sim}{\theta}_{ij}, \underset{\sim}{\theta}_{rj}^T) = \underset{\sim}{\Sigma}_b, \qquad i \neq r , \tag{4c}$$

$$\text{cov}(\underset{\sim}{\theta}_{ij}, \underset{\sim}{\theta}_{ij}^T) = \underset{\sim}{\Sigma}_a + \underset{\sim}{\Sigma}_b + \underset{\sim}{\Sigma}_c , \tag{4d}$$

and we alternatively express our prior structure by saying that the vectors $\underset{\sim}{\theta}_{ij}$ have means $\underset{\sim}{\mu}$ and dispersions described by equations (4). In the language of the general linear model developed by Lindley and Smith (1972), the second stage has a dispersion matrix $\underset{\sim}{C}_2$ whose elements are described by (4): these elements themselves being matrices. To find our estimates of $\underset{\sim}{\theta}_{ij}$, we use the general theory in the way described in Appendix 1 of L . To do this, we have first to invert $\underset{\sim}{C}_2$. This is most easily done by solving the linear equations in z, $\underset{\sim}{C}_2\underset{\sim}{z} = \underset{\sim}{a}$. These equations* are easily seen to be [cf. (1.3) of L]

$$\underset{\sim}{\Sigma}_c\underset{\sim}{z}_{ij} + n\underset{\sim}{\Sigma}_a\underset{\sim}{z}_{i\cdot} + m\underset{\sim}{\Sigma}_b\underset{\sim}{z}_{\cdot j} = \underset{\sim}{a}_{ij} \quad . \tag{5}$$

For any array u_{ij}, vector or scalar, write

$$u'_{ij} = u_{ij} - u_{i\cdot} - u_{\cdot j} + u_{\cdot\cdot} \; ,$$

$$u'_{i\cdot} = u_{i\cdot} - u_{\cdot\cdot} \; ,$$

$$u'_{\cdot j} = u_{\cdot j} - u_{\cdot\cdot} \; ,$$

$$u'_{\cdot\cdot} = u_{\cdot\cdot} \; .$$

Then, (5) may be rewritten

$$\underset{\sim}{V}_o\underset{\sim}{z}'_{ij} + \underset{\sim}{V}_n\underset{\sim}{z}'_{i\cdot} + \underset{\sim}{V}_m\underset{\sim}{z}'_{\cdot j} + \underset{\sim}{V}_{mn}\underset{\sim}{z}'_{\cdot\cdot} = \underset{\sim}{a}_{ij} \tag{6}$$

where

$$\left.\begin{array}{l} \underset{\sim}{V}_o = \underset{\sim}{\Sigma}_c, \; \underset{\sim}{V}_n = \underset{\sim}{\Sigma}_c + n\underset{\sim}{\Sigma}_a, \; \underset{\sim}{V}_m = \underset{\sim}{\Sigma}_c + m\underset{\sim}{\Sigma}_b \; , \\ \text{and} \\ \underset{\sim}{V}_{mn} = \underset{\sim}{\Sigma}_c + n\underset{\sim}{\Sigma}_a + m\underset{\sim}{\Sigma}_b \; . \end{array}\right\} \tag{7}$$

*Notice that in writing out these and similar equations in mnp unknowns, we have written them as mn equations each involving p-vectors. In this way, their structure is most easily understood and related to the scalar forms in L .

The summation of (6) over all i, j immediately gives $\underset{\sim}{V}_{mn}\underset{\sim}{z}'_{..} = \underset{\sim}{a}_{..}$; over j alone gives $\underset{\sim}{V}_n\underset{\sim}{z}'_{i.} = \underset{\sim}{a}'_{i.}$; over i alone gives $\underset{\sim}{V}_m\underset{\sim}{z}'_{.j} = \underset{\sim}{a}'_{.j}$, and hence, substituting in (6) $\underset{\sim}{V}_o\underset{\sim}{z}'_{ij} = \underset{\sim}{a}'_{ij}$. The solution to $\underset{\sim}{C}_2\underset{\sim}{z} = \underset{\sim}{a}$ is, therefore,*

$$\underset{\sim}{z}_{ij} = \underset{\sim}{V}_o^{-1}\underset{\sim}{a}'_{ij} + \underset{\sim}{V}_n^{-1}\underset{\sim}{a}'_{i.} + \underset{\sim}{V}_m^{-1}\underset{\sim}{a}'_{.j} + \underset{\sim}{V}_{mn}^{-1}\underset{\sim}{a}'_{..} \quad . \tag{8}$$

The coefficients of $\underset{\sim}{a}_{ij}$ (<u>not</u> $\underset{\sim}{a}'_{ij}$) on the right-hand side of this equation are the elements in $\underset{\sim}{C}_2^{-1}$. If the elements of this matrix are termed the <u>coprecisions</u>, and abbreviated cop, we have [cf. equations (4) above]

$$\text{cop}(\underset{\sim}{\theta}_{ij}, \underset{\sim}{\theta}_{rs}^T) = \underset{\sim}{H}_o, \qquad i \neq r,\ j \neq s\ , \tag{9a}$$

$$\text{cop}(\underset{\sim}{\theta}_{ij}, \underset{\sim}{\theta}_{is}^T) = \underset{\sim}{H}_a + \underset{\sim}{H}_o, \qquad j \neq s\ , \tag{9b}$$

$$\text{cop}(\underset{\sim}{\theta}_{ij}, \underset{\sim}{\theta}_{rj}^T) = \underset{\sim}{H}_b + \underset{\sim}{H}_o, \qquad i \neq r\ , \tag{9c}$$

$$\text{cop}(\underset{\sim}{\theta}_{ij}, \underset{\sim}{\theta}_{ij}^T) = \underset{\sim}{H}_c + \underset{\sim}{H}_a + \underset{\sim}{H}_b + \underset{\sim}{H}_o \tag{9d}$$

where

$$mn\underset{\sim}{H}_o = \underset{\sim}{V}_o^{-1} - \underset{\sim}{V}_n^{-1} - \underset{\sim}{V}_m^{-1} + \underset{\sim}{V}_{mn}^{-1}\ , \tag{10a}$$

$$n\underset{\sim}{H}_a = -\underset{\sim}{V}_o^{-1} + \underset{\sim}{V}_n^{-1}\ , \tag{10b}$$

$$m\underset{\sim}{H}_b = -\underset{\sim}{V}_o^{-1} + \underset{\sim}{V}_m^{-1}\ , \tag{10c}$$

and

$$\underset{\sim}{H}_c = \underset{\sim}{V}_o^{-1}\ . \tag{10d}$$

Having found $\underset{\sim}{C}_2^{-1}$, we need to evaluate

$$\underset{\sim}{C}_2^{-1}\underset{\sim}{A}_2(\underset{\sim}{A}_2^T\underset{\sim}{C}_2^{-1}\underset{\sim}{A}_2)^{-1}\underset{\sim}{A}_2^T\underset{\sim}{C}_2^{-1} \tag{11}$$

*The reader can check that with p = 1, this agrees with the result (1.10) in L . The new form given here is more convenient in the matrix case.

where $\underset{\sim}{A}_2$ is a matrix of mnp rows and p columns all, mn, of whose p x p submatrices are unit matrices. This last fact follows since every θ_{ij} has the same expectation, μ . It is easy to verify that the column totals for the submatrices in $\underset{\sim}{C}_2^{-1}$ are all $\underset{\sim}{V}_{mn}^{-1}$, and that from this it follows that (11) is a square matrix of dimension mnp all of whose p x p submatrices are $(mn\underset{\sim}{V}_{mn})^{-1}$. This completes the calculations needed at the second stage of the model.

The first stage of the model is described by (2), and in the notation of the general linear model, this is given in terms of matrix $\underset{\sim}{A}_1$ of a form with matrices $\underset{\sim}{X}_{ij}$ down the "diagonal" and zeros elsewhere, the matrix $\underset{\sim}{X}_{ij}$ having r_{ij} rows and p columns with typical (k, s) element x_{ijks} . The corresponding dispersion matrix, $\underset{\sim}{C}_1$, is simply σ^2 times a unit matrix, and so the matrix $\underset{\sim}{A}_1^T\underset{\sim}{C}_1^{-1}\underset{\sim}{A}_1$ that occurs in the usual normal equations has diagonal submatrices $\underset{\sim}{X}_{ij}^T\underset{\sim}{X}_{ij}\sigma^{-2}$, all of size p x p, and zeros elsewhere. $\underset{\sim}{X}_{ij}^T\underset{\sim}{X}_{ij}$ is simply the matrix of sums of squares and products of the x-values occurring in cell (i, j); the typical (s, t) element being $\sum_k x_{ijks}x_{ijkt}$. Similarly, on the other side of the normal equation, we shall have terms $\sum_k y_{ijk}x_{ijks}$. These produce a column vector consisting of subvectors, a typical one of which is $\underset{\sim}{X}_{ij}^T\underset{\sim}{y}_{ij}\sigma^{-2}$ where $\underset{\sim}{y}_{ij}$ is the vector of elements y_{ijk} .

We are now in a position to write down the equations satisfied by the modal estimates of θ_{ijs} assuming $\underset{\sim}{\Sigma}_a$, $\underset{\sim}{\Sigma}_b$, $\underset{\sim}{\Sigma}_c$, and σ^2 to be known [cf. equations (3.1) of L]. In writing these out, we should distinguish between θ_{ijs} and the estimate θ^*_{ijs}, say, but to avoid even more complicated notation, the asterisk will be omitted. The equations* are

$$(\underset{\sim}{H}_c + \underset{\sim}{X}_{ij}^T\underset{\sim}{X}_{ij}\sigma^{-2})\underset{\sim}{\theta}_{ij} + n\underset{\sim}{H}_a\underset{\sim}{\theta}_{i\cdot} + m\underset{\sim}{H}_b\underset{\sim}{\theta}_{\cdot j}$$

$$+ [mn\underset{\sim}{H}_o - \underset{\sim}{V}_{mn}^{-1}]\underset{\sim}{\theta}_{\cdot\cdot} = \underset{\sim}{X}_{ij}^T\underset{\sim}{y}_{ij}\sigma^{-2} \quad . \qquad (12)$$

*In the notation of Appendix 7 of L, they are $\underset{\sim}{D}^{-1}\underset{\sim}{\theta}_1 = \underset{\sim}{d}$.

[The corresponding equations in the scalar case are (3.1) in L .] This is the main result of the paper since it provides the adjusted estimates of the regression coefficients that replace the usual least-squares estimates. The latter would be the solution of the normal equations

$$X_{ij}^T X_{ij} \theta_{ij} = X_{ij}^T y_{ij}$$

obtained from (12) by omitting all the H-matrices and V_{mn} . In the normal equations, the regression coefficients in cell (i, j) are not involved with the coefficients in any other cell, whereas all occur in (12). The basic equations may be rewritten in terms of the original matrices Σ_a, Σ_b, and Σ_c using equations (10) and (7). The result is

$$(\Sigma_c^{-1} + X_{ij}^T X_{ij} \sigma^{-2})\theta_{ij} + \left[(\Sigma_c + n\Sigma_a)^{-1} - \Sigma_c^{-1}\right]\theta_{i.}$$

$$+ \left[(\Sigma_c + m\Sigma_b)^{-1} - \Sigma_c^{-1}\right]\theta_{.j} + \left[\Sigma_c^{-1} - (\Sigma_c + n\Sigma_a)^{-1} - (\Sigma_c + m\Sigma_b)^{-1}\right]\theta_{..}$$

$$= X_{ij}^T y_{ij} \sigma^{-2} \quad . \qquad (13)$$

To complete the analysis, it only remains to provide equations for the estimation of Σ_a, Σ_b, Σ_c, and σ^2 . To do this, we return to the structure of the vectors of regression coefficients, θ_{ij}, described in (3), resulting in the covariance values given in (4). In the last of these results, (4d), the three dispersion matrices occur in combination, and any attempt to base an estimation procedure on this has difficulty in separating the components. This can be seen again in (13) where, for example, Σ_a always occurs in association with Σ_c . To circumvent this difficulty, we redescribe the model in terms of an extra stage, and suppose, firstly that $\theta_{ij} \sim N_p(\alpha_i + \beta_j, \Sigma_c)$ and then that both $\alpha_i \sim N_p(\mu_a, \Sigma_a)$ and $\beta_j \sim N_p(\mu_b, \Sigma_b)$ with vague prior knowledge of μ_a and μ_b .

Clearly, this is an alternative way of expressing (3). With this model, we can write down the joint probability distribution of all the quantities in the problem. After integration over the diffuse priors for μ_a and μ_b, this is easily seen to be proportional to (where $R = \Sigma r_{ij}$)

$$\sigma^{-R}\exp\left[-\tfrac{1}{2}\sum_{i,j,k}(y_{ijk} - \underset{\sim}{\theta}_{ij}^T \underset{\sim}{x}_{ijk})^2\right]$$

$$\times \left|\underset{\sim}{\Sigma}_c\right|^{-\frac{1}{2}mn}\exp\left[-\tfrac{1}{2}\sum_{i,j}(\underset{\sim}{\theta}_{ij} - \underset{\sim}{\alpha}_i - \underset{\sim}{\beta}_j)^T \underset{\sim}{\Sigma}_c^{-1}(\underset{\sim}{\theta}_{ij} - \underset{\sim}{\alpha}_i - \underset{\sim}{\beta}_j)\right]$$

$$\times \left|\underset{\sim}{\Sigma}_a\right|^{-\frac{1}{2}(m-1)}\exp\left[-\tfrac{1}{2}\sum_{i}(\underset{\sim}{\alpha}_i - \underset{\sim}{\alpha}_.)^T \underset{\sim}{\Sigma}_a^{-1}(\underset{\sim}{\alpha}_i - \underset{\sim}{\alpha}_.)\right]$$

$$\times \left|\underset{\sim}{\Sigma}_b\right|^{-\frac{1}{2}(n-1)}\exp\left[-\tfrac{1}{2}\sum_{j}(\underset{\sim}{\beta}_j - \underset{\sim}{\beta}_.)^T \underset{\sim}{\Sigma}_b^{-1}(\underset{\sim}{\beta}_j - \underset{\sim}{\beta}_.)\right] . \qquad (14)$$

In explanation, the first line describes the likelihood of the data, the second the distribution of the regression coefficients given the row and column values (essentially, this is the interaction term), and the last two lines provide the distributions of the row and column values, after removal of their means by integration.

This is, by Bayes theorem, proportional to the posterior distribution of $(\underset{\sim}{\theta}_{ij})$, $(\underset{\sim}{\alpha}_i)$, and $(\underset{\sim}{\beta}_j)$--if the dispersion are known. The modal values of the $\underset{\sim}{\theta}$'s have been found as the solutions to equations (12) or (13). We now determine the similar estimates of the $\underset{\sim}{\alpha}$'s and $\underset{\sim}{\beta}$'s. To do this, differentiate the logarithm of (14) with respect to $\underset{\sim}{\alpha}_i$ and $\underset{\sim}{\beta}_j$ with the results

$$\left.\begin{aligned} & n\underset{\sim}{\Sigma}_c^{-1}(\underset{\sim}{\theta}_{i.} - \underset{\sim}{\alpha}_i - \underset{\sim}{\beta}_.) + \underset{\sim}{\Sigma}_a^{-1}(\underset{\sim}{\alpha}_i - \underset{\sim}{\alpha}_.) = 0 \\ \text{and} \\ & m\underset{\sim}{\Sigma}_c^{-1}(\underset{\sim}{\theta}_{.j} - \underset{\sim}{\alpha}_. - \underset{\sim}{\beta}_j) + \underset{\sim}{\Sigma}_b^{-1}(\underset{\sim}{\beta}_j - \underset{\sim}{\beta}_.) = 0 \quad . \end{aligned}\right\} \qquad (15)$$

Summation of the first (say) set over i gives $\underset{\sim}{\alpha}_. + \underset{\sim}{\beta}_. = \underset{\sim}{\theta}_{..}$ and, hence, eliminating $\beta_.$ from a member of that set, we have

$$n\underset{\sim}{\Sigma}_c^{-1}(\underset{\sim}{\theta}_{i.} - \underset{\sim}{\theta}_{..}) = (n\underset{\sim}{\Sigma}_c^{-1} + \underset{\sim}{\Sigma}_a^{-1})(\underset{\sim}{\alpha}_i - \underset{\sim}{\alpha}_.)$$

or

$$\left.\begin{aligned} (\underset{\sim}{\alpha}_i - \underset{\sim}{\alpha}_.) &= (\underset{\sim}{\Sigma}_c + n\underset{\sim}{\Sigma}_a)^{-1} n\underset{\sim}{\Sigma}_a(\underset{\sim}{\theta}_{i.} - \underset{\sim}{\theta}_{..}) \quad . \\ \text{Similarly,} \qquad & \\ (\underset{\sim}{\beta}_j - \underset{\sim}{\beta}_.) &= (\underset{\sim}{\Sigma}_c + m\underset{\sim}{\Sigma}_b)^{-1} m\underset{\sim}{\Sigma}_b(\underset{\sim}{\theta}_{.j} - \underset{\sim}{\theta}_{..}) \quad . \end{aligned}\right\} \tag{16}$$

Since the estimates of the $\underset{\sim}{\theta}$'s are known, these equations determine the estimates of the $\underset{\sim}{\alpha}$'s and $\underset{\sim}{\beta}$'s.

To complete the analysis, we have to insert prior distribution for σ^2, $\underset{\sim}{\Sigma}_a$, $\underset{\sim}{\Sigma}_b$, and $\underset{\sim}{\Sigma}_c$. We suppose that these are independent, and for σ^2 there is no objection to using the usual vague prior proportional to σ^{-2} . With the three dispersion matrices, one cannot be so cavalier. In default of any better idea, we assume that $\underset{\sim}{\Sigma}_u^{-1}$ has a Wishart distribution with degrees of freedom ν_u and matrix $\underset{\sim}{\Delta}_u(u = a, b, c)$. Specifically, $\underset{\sim}{\Sigma}_a^{-1}$ has density proportional to

$$\left|\underset{\sim}{\Delta}_a\right|^{\frac{1}{2}\nu_a}\left|\underset{\sim}{\Sigma}_a^{-1}\right|^{\frac{1}{2}(\nu_a-p-1)} \exp\left[-\tfrac{1}{2}\nu_a \operatorname{tr}(\underset{\sim}{\Delta}_a\underset{\sim}{\Sigma}_a^{-1})\right] \tag{17}$$

with $\underset{\sim}{\Sigma}_b^{-1}$ and $\underset{\sim}{\Sigma}_c^{-1}$, similarly. The choice of values for the ν's and $\underset{\sim}{\Delta}$'s will be discussed below.

The procedure now is to combine the prior distribution, exemplified by (17), with the other terms in (14) so obtaining the joint posterior distribution for which the modal values can be determined by differentiation and equating the results to zero. The analysis closely parallels that given by Lindley and Smith (1972, Section 5.2) and the results are that

$$\sigma^2 = \sum_{i,j,k} (y_{ijk} - \theta^T_{ij} x_{ijk})^2 \Big/ (R + 2) \quad , \tag{18a}$$

$$\Sigma_c = \left[\nu_c \Delta_c + \sum_{i,j} (\theta_{ij} - \alpha_i - \beta_j)(\theta_{ij} - \alpha \qquad) \right] \Big/ (mn + \nu_c - p - 1) \ , \tag{18b}$$

$$\Sigma_b = \left[\nu_b \Delta_b + \sum_i (\alpha_i - \alpha_.)(\alpha_i - \alpha_.)^T \right] \Big/ (m + \nu_b - p - 2) \ , \tag{18c}$$

and

$$\Sigma_a = \left[\nu_a \Delta_a + \sum_j (\beta_j - \beta_.)(\beta_j - \beta_.)^T \right] \Big/ (n + \nu_a - p - 2) \ . \tag{18d}$$

(In this set of equations, as in (12), the estimates--for example θ^*_{ij}--should strictly replace the values like θ_{ij} written there.)

To perform the analysis, values have to be specified for the ν's and Δ's. Since ν_u measures ones prior precision about Σ_u, it is natural to take ν_u as small as possible. The least value compatible with the convergence of the prior distributions in (17) is $\nu_u = p$, and we suggest taking this value. There remains the choice of Δ_u .

In the form we have written (17), Δ_a^{-1} is the expectation of Σ_a^{-1} so that Δ_a can be thought of as ones prior opinion of the dispersion of the α's, and similarly, Δ_b for the β's and Δ_c for the γ's. One suggestion is to rescale the regression variables $x_1, x_2, \ldots, x_p$ so that in the new scale, the prior opinions about scatter of all the α_i are the same, then Δ_a will have a constant diagonal. However, once this scaling has been done, we are not free to make a new scaling so that Δ_b (and Δ_c) have constant diagonals. Hence, in general, this device is not available. Nevertheless, we conjecture that in many cases a common scaling might be reasonable. It would, for example, be a fairly sophisticated form of prior knowledge that thought that α_{i7} (the last suffix referring to x_7) was more dispersed than α_{i5}, whereas β_{j7} was <u>less</u> dispersed than β_{j5} : in other words, the rows would have more effect on the regression coefficients for the seventh variable than on the fifth, whereas the situation would be reversed for the columns. The simplest assumption is, therefore, to suppose $\underset{\sim}{\Delta}_a$, $\underset{\sim}{\Delta}_b$, and $\underset{\sim}{\Delta}_c$, in a suitable scaling of the x's, all have constant diagonal entries, though possibly varying with the matrices--the interaction, for example, might, apriori, be judged smaller than the main effects.

Having settled on the diagonal elements in $\underset{\sim}{\Delta}_u$, it remains to consider the off-diagonal values, or effectively, the correlations. There would seem to be no real objection to putting these zero, except in one important case. As explained above, the first predictor variable will typically be a constant to allow for the regression surface not passing through the origin. In this case, θ_{ij1} has rather a different status from the other θ's, and in particular, it is changed substantially by altering the <u>origin</u> of any predictor variable. (In the last paragraph, changes of <u>scale</u> were under discussion.) Jackson has suggested changing the origin of each genuine (that is, apart from the constant x_1) predictor variable so that, apriori, the correlation between α_{i1} and α_{is} $(s \neq 1)$ is zero. But again, there is no necessity for this origin being the same for the $\underset{\sim}{\alpha}$'s as for the $\underset{\sim}{\beta}$'s or $\underset{\sim}{\gamma}$'s. If we suppose, on the lines of the argument in the last paragraph, that they are, then each of $\underset{\sim}{\Delta}_a$, $\underset{\sim}{\Delta}_b$, and $\underset{\sim}{\Delta}_c$ may be taken to be multiples of the unit matrix.

This completes the theory of the two-way regression analysis. Essentially, we have two groups of equations, (13) and (18), to solve for $(\underset{\sim}{\theta}_{ij})$, σ^2, $\underset{\sim}{\Sigma}_a$, $\underset{\sim}{\Sigma}_b$, and $\underset{\sim}{\Sigma}_c$. [The $\underset{\sim}{\alpha}$'s and $\underset{\sim}{\beta}$'s that occur in (18c) and (18d) can be eliminated using (16).] We have to feed in ν_a, ν_b, and ν_c--we suggest putting them all equal to p--$\underset{\sim}{\Delta}_a$, $\underset{\sim}{\Delta}_b$, and $\underset{\sim}{\Delta}_c$--we suggest, after suitable changes of origin and scale for the genuine regression variables, putting these all equal to a multiple of unit matrix, say $\underset{\sim}{\Delta}_u = \delta_u \underset{\sim}{I}$. The data enter through the sufficient statistics $\underset{\sim}{X}_{ij}^T \underset{\sim}{X}_{ij}$, $\underset{\sim}{X}_{ij}^T \underset{\sim}{y}_{ij}$, and the sum of squares in (18a). Equations (13) and (18) are fairly involved and decidely non-linear. They, therefore, raise several problems in numerical analysis when we go to solve them. We only offer a few comments on a possible mode of solution.

Notice that any computational procedure ought to output the inverse of the matrix on the left-hand side of (13) since this provides the dispersion matrix of the posterior distribution of $(\underset{\sim}{\theta}_{ij})$, exactly if the dispersions are known, approximately otherwise. We don't provide a similar dispersion matrix for the $\underset{\sim}{\Sigma}$'s since this posterior is not well-understood. Indeed, we suspect that the estimates of the $\underset{\sim}{\Sigma}$'s may not be too good, nevertheless, we believe that any errors here will not seriously affect the estimation of the quantities of primary interest, the $\underset{\sim}{\theta}$'s .

One possible way to solve these equations is to start with the least-squares estimates for $\underset{\sim}{\theta}_{ij}$, to identify $\underset{\sim}{\alpha}_i$ with $\underset{\sim}{\theta}_{i.}$ and $\underset{\sim}{\beta}_j$ with $\underset{\sim}{\theta}_{.j}$, (16)--this amounts to putting $\underset{\sim}{\Sigma}_c = 0$--and then estimating σ^2, $\underset{\sim}{\Sigma}_a$, $\underset{\sim}{\Sigma}_b$, and $\underset{\sim}{\Sigma}_c$ from (18). With these values, solve (13) for $\underset{\sim}{\theta}_{ij}$ and obtain the $\underset{\sim}{\alpha}$'s and $\underset{\sim}{\beta}$'s from (16). These new values can be inserted in (18) again and the cycle repeated. Another possibility is, instead of starting with the least-squares values, guess a value for σ^2 and solve (18) with this value and $\underset{\sim}{\Sigma}_u = \underset{\sim}{\Delta}_u$ (u = a, b, c). From these values, the $\underset{\sim}{\alpha}$'s and $\underset{\sim}{\beta}$'s could be found with the same original guesses, and then the iterative procedure suggested above repeated.

References

Jackson, P. H., Novick, M. R., & Thayer, D. T. "Estimating regressions in m groups." The British Journal of Mathematical and Statistical Psychology, 1971, 24, 129-153.

Jones, P. K., & Novick, M. R. "Implementation of a Bayesian system for prediction in m groups." ACT Technical Bulletin No. 6. Iowa City, Iowa: The American College Testing Program, 1972.

Lindley, D. V. "A Bayesian solution for some educational prediction problems, II." Research Bulletin 69-91. Princeton, N.J.: Educational Testing Service, 1969.

Lindley, D. V. "A Bayesian solution for some educational prediction problems, III." Research Bulletin 70-33. Princeton, N.J.: Educational Testing Service, 1970.

Lindley, D. V. "A Bayesian solution for two-way analysis of variance." ACT Technical Bulletin No. 8. Iowa City, Iowa: The American College Testing Program, 1972, in press.

Lindley, D. V., & Smith, A. F. M. "Bayes estimates for the linear model." Journal of the Royal Statistical Society, Series B, 1972, 34, 1-41.

Novick, M. R., Jones, P. K., & Cole, N. S. "Predictions of performance in career education." Research Report No. 55. Iowa City, Iowa: The American College Testing Program, 1972, in press.